

12. A computer-readable memory containing a computer program that is executable by a computer to perform the process recited in Claim 9.

13. A computer-readable medium having computer-executable instructions for use in detecting when a person in a synchronized audio video clip is speaking, said computer executable instructions comprising:

5           inputting one or more captured video and synchronized audio clips, segmenting said audio and video clips to remove portions of said video and synchronized audio clips not needed in determining if a speaker in the  
10          captured video and synchronized audio clips is speaking;

             extracting audio and video features in said captured video and synchronized audio clips to be used in determining if a speaker in the captured; and wherein an audio feature is the energy over an audio frame and wherein said video feature is the openness of a person's mouth;

15          training a Time Delay Neural Network to determine when a person is speaking using said extracted audio and video features.

*Rule 1.176*  
<sup>14</sup>  
15. The computer-readable medium of Claim 14 wherein the instruction for training a Time Delay Neural Network further comprises a sub-instruction for  
20          correlating said audio features and video features to determine when a person is speaking.